UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR

ANOMALY DETECTION IN VIDEO SEQUENCES

Luis Alberto Caro Campos
Supervisor: Juan Carlos San Miguel Avedillo

-TRABAJO FIN DE MASTER-

Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Octubre 2013

ANOMALY DETECTION IN VIDEO SEQUENCES

Luis Alberto Caro Campos
Supervisor: Juan Carlos San Miguel Avedillo
email: {Luis.Caro, JuanCarlos.SanMiguel}@uam.es

Video Processing and Understanding Lab
Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Octubre 2013

Trabajo parcialmente financiado por el Gobierno de España bajo el proyecto TEC2011-25995 (EventVideo)

Abstract

In this work, a comprehensive study of an existing anomaly detection framework has been carried out. After identifying current challenges in the field of anomaly detection in video sequences, an existing framework has been selected for its implementation and evaluation. A set of video sequences containing anomalies from common surveillance scenarios have been selected from publicly available datasets. The system has been evaluated on these video sequences in order to identify existing shortcomings. Improvements to the original algorithm have been proposed in order to address the observed limitations. Finally, the performance of the proposed changes has been evaluated on the same video sequences for comparison.

Acknowledgements

Ante todo, me gustaría agradecer a Chema y a Juan Carlos por su inestimable ayuda, disponibilidad y paciencia a lo largo de los últimos dos años.

A todos los miembros de VPU-Lab, con quienes me he sentido como en casa desde el primer día.

A mi familia, ya que sin su apoyo y su cariño no habría llegado hasta aquí. Porque juntos somos capaces de superar cualquier adversidad.

A mis amigos, por estar ahí sin importar la distancia.

Por último, me gustaría dedicar este trabajo a mis hermanos, *Carlos y Agustín*.

Luis Caro Campos

Octubre de 2013

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Nowadays, there is a growing demand for automated surveillance systems due to heightened security concerns. Technological advances and reduced costs have led to an accelerated deployment of surveillance cameras, both in public and private facilities. Traditionally, the monitoring task is performed by human operators who have to visually inspect video feeds from multiple cameras at the same time. However, it has been shown that even dedicated personnel are affected by a diminished visual attention after long periods of staring at monitor screens [1]. This hinders their ability to detect and react to potential threats in real-time [2], turning current surveillance systems into mere recording devices that are used only for post-event video forensics [3]. For these reasons, real-time event recognition in video surveillance has gained interested as a research topic over the last decade. Automated video surveillance can assist human operators in identifying potential threads, as well as issuing appropriate response where needed.

In this context, human activity recognition has been widely studied in the literature. Most approaches in this field explicitly model certain events a priori, and therefore their application is limited to the detection of these events, usually in controlled scenarios. Examples of event detection include abandoned object detection, or trespassing of forbidden areas. More recently, there has been increased focus on anomaly detection without explicit modeling. Behind this surge of interest, is the fact that events of interest in video surveillance scenarios are sparse and hard to predict, and therefore it is impossible to train a system to cover all possible cases of anomalous events.

Common to these techniques is the underlying assumption that anomalous events

are characterized by their low frequency of occurence when compared to normal events.

1.2 Objectives

The main objective of this Masters Thesis is to develop a framework for the detection of anomalous events in video sequences and to evaluate its performance on scenarios that pose challenges for current approaches in the literature. After a comprehensive study of the literature, one anomaly detection algorithms will be selected for its implementation and improvements will be proposed. In particular, contextual and behavioural methods will be taken into consideration for their evaluation. The following sub-goals are defined:

- Compilation of a video data set for anomaly detection: this set should be comprehensive and cover a wide variety of scenarios in terms of types of events, objects, object clutter and crowds.
- Selection of a base algorithm for its implementation: an anomaly detection algorithm will be selected and implemented.
- Improvements of base algorithm for challenging scenarios: after a preliminary evaluation of the algorithm on the data set, key challenging areas will be identified, and improvements will be explored.

1.3 Document Structure

This document is structured as follows:

- Chapter 1. This chapter presents the motivation of objectives of this Masters Thesis.
- Chapter 2. In this chapter, an overview of the literature related to the field of anomaly detection in video sequences is presented
- Chapter 3. This chapter describes the selected algorithm for its implementation and evaluation
- Chapter 4. In this chapter, an evaluation of the performance of the base system on common surveillance scenarios is presented.
- Chapter 5. This chapter describes the proposed enhancements to the base system, in order to provide robustness against challenging scenarios.

- Chapter 6. This chapter presents the evaluation of the proposed modifications to the base system on different scenarios.
- Chapter 7. This chapter summarizes the main contributions of this work, discusses the results obtained and provides suggestions for future lines of research.

Chapter 2

State Of The Art

2.1 Introduction

Recognition and understand of human activity in video have gained considerable research attention due to the potential application in various domains [4], such as video surveillance and monitoring, human-computer interfaces, content-based video analysis and behavioural biometrics. In video surveillance, the main objective is to be able to detect events of interest to aid security personnel. In the literature, we can find works that perform event detection by learning patterns for specific events depending on the domain. Many of the works in this area take a traditional pattern recognition approach by explicitly modelling the events of interest based on a priori knowledge. Examples include parking in restricted areas [5], detection of abandoned objects [6], and suspicious interactions with objects [7].

More recently, there has been a paradigm shift [8] towards detection of anomalous events. Intuitively, an anomaly is a pattern that does not follow expected normal behaviour in a given context [9]. The anomaly detection problem has been studied in the literature in very diverse application domains [9]. For example, anomalous traffic patterns in a computer network may indicate an intrusion attempt which could be thwarted if measures are taken in time; in medical imaging, detecting anomalies could aid in the diagnosis of certain conditions; anomalous credit card transactions can also indicate fraudulent activity and prompt banks to block them to prevent financial loss.

In video surveillance, pattern recognition techniques are often the desired approach to perform event detection and behaviour understanding, by modelling anomalous events from training data. Detection is performed by finding patterns in new observations that conform to the previously obtained model. In contrast, in anomalous event detection, a model of normal behaviour is developed statistically and

anomalies are detected by finding patterns that deviate from the model.

By specifically targeting events of interest, the more traditional approaches are able to provide high level descriptors of events occurring in the scene. These techniques construct models from training data that contain instances of the targeted anomalies, and attempt to classify as anomalies new unseen instances. Their main limitation, however, is their inability to cope with unknown behaviour, and are generally only applicable in certain controlled scenarios. Anomaly detection techniques, in contrast, are able detect arbitrary anomalies that differ from a previously obtained model of normality. While this approach broadens the amount of events that can be detected, it still poses significant challenges depending on how normality is defined. Additionally, these techniques are unable describe "what" has occurred, and would require further stages of analysis to provide higher level descriptors.

The rest of this chapter is structured as follows. Sections 2.2 and 2.3 cover the definition of an anomaly and a discussion on the different ways in which anomalies that can be defined. In section 2.4, a discusion on current challenges in anomaly detection techniques. In section 2.5, a brief discussion on different types of features is introduced. In section 2.6 a review of approaches that rely on object trajectories is presented, while alternative approaches are reviewed in section 2.7. Section 2.8 discusses current evaluation frameworks for validating existing approaches. Finally, concluding remarks are presented in section 2.9.

2.2 What is an anomaly

In spite of the diversity of solutions and applications, there is a lack of agreement on how anomalies are defined. In the literature, anomalies have been referred to as “unusual events” [10], “anomalous events” [11], “abnormality” [12], “suspicious activities” or “irregularities” [13].

In broad terms, we can define an anomaly as an observation that does not follow expected normal behaviour [9]. For video sequences, anomalous events can be seen as motions or sequence of motions that stand out in their surrounding context in space and time [8]. This enables a statistical treatment of anomaly detection, by considering anomalies as events of low probability with respect to a probabilistic model of normal behaviour. [9][14].

This definition has certain implications that limit which anomalies can detected. Firstly, it makes anomalies dependent on a given context. An event that is anomalous at a certain moment, may be perfectly normal at other times. Such is the case of traffic interactions, in which certain actions are only allowed given certain conditions

like the state of traffic lights. Secondly, the anomalous events that can be detected are directly limited by the features and the scale at which normality is defined[14]. E.g., an event that is anomalous at a certain scale may be perfectly normal at a different scale.

As a result of the different ways in which anomalous events can be defined, very diverse approaches can be found in the literature.

2.3 Types of anomalies

Depending on the levels of spatio-temporal context and which anomalous events are modelled, we can broadly distinguish between the following different types of anomalies [9].

Point anomalies indicate that the values of extracted features at a specific location deviate significantly from what is considered normal. Therefore, these anomalies do not take into account past values or the information given by nearby objects or points. If one models the normal velocities of moving objects at all locations in the scene, any object that displays a velocity that does not fit the model can be considered an anomaly. This includes, for example, detecting motion of objects at unusual locations.

Contextual anomalies consider information from the temporal context (the sequence of events), or the spatial context (nearby objects). Anomalies that take into account the temporal context, also called sequential anomalies, analyse irregularities in the temporal sequence of a given extracted feature. For example, in traffic sequences, a car making an illegal turn at an intersection may display "normal" velocity as it passes through it, as it trajectory will partially overlap different normal traffic paths. However, the trajectory itself is anomalous as it does deviate from the predefined path from that direction. For anomalies in the spatial context, information from nearby objects is taken into account. For example, the authors in [15] analyse the avoidance strategies of nearby people to detect anomalies in walking paths. In [11], the authors consider co-occurrence anomalies, by detecting paris of events that do not usually occur simultaneously.

These distinctions highlight the fact that anomalies are heavily dependent on the given context, and can be arbitrarily complex depending on the features extracted. As a result, it becomes difficult to directly compare diverse solutions found in the

literature. Depending on the capabilities of the algorithms, different sets of anomalies can be found on the same datasets.

2.4 Challenges

While anomalous events are easy to define intuitively, there are a number of factors that pose challenges to anomaly detection techniques:

- The definition of anomaly is heavily dependant on how normality is modelled and which features are extracted. In particular, context, features, and the scale at which features are extracted will ultimately determine which anomalies can be detected.

- A non stationary context may alter normality at different times in a given scenario. A robust detection system should be able to adapt to changing dynamics to account for these changes.

- Anomalous events are generally infrequent, sparse, and unpredictable [14]. This makes the examples found in training sequences limited in number. In particular, validation of techniques becomes a challenge if the number of anomalous events are insufficient.

2.5 Features

According to [8], we can distinguish between pixel based abstractions (features extracted at the pixel level) and object-based abstractions (features associated with an object or blob).

Among pixel-based abstractions, we find approaches that capture spatio-temporal features such as pixel change frequency and pixel change retainment [16], filling ratio of foreground pixels [17], histogram of pixel change frequency [18], gradient magnitude [13], accumulation of pixel differences [19]. Motion features are also common, extracted by optical flow techniques at the pixel level [20][21].

Among object-based abstractions, they can either be derived from appearance features or motion features. Among appearance features, we can include blob size[22][21] [23][17] and texture [21]. Motion features derived from object tracking are widely popular. Object tracking produces trajectories as a sequence of object location over time. From these, different features such as velocity/speed [24][11], and moving direction or orientation [22][25][20] can be extracted.

Among existing techniques, we can make a broad distinction between approaches that rely on extracted object trajectories, and those that do not [8][25]. The former imply preprocessing modules for object segmentation and tracking, whereas the later rely on other object features, or pixel-based abstractions.

2.6 Approaches based on object trajectories

Several approaches found in the literature for anomaly detection in video sequences employ information extracted from object trajectories, i.e., the temporal sequence of locations of a given object in the scene. Therefore, these techniques require a pre-processing stage in which moving object detection and object tracking are performed. Generally, background subtraction is employed for moving object detection and existing tracking techniques can be applied to extract object trajectories.

The main advantage of trajectory-based techniques is the possibility of constructing models in a fully unsupervised manner, i.e., labeling training data is not required. By considering anomalies as events with low frequency, clustering methods can be applied to discard outliers [11]. This is done by clustering trajectory paths to model "normal" trajectories. Anomalies are then detected by computing the distance of new unseen trajectories to existing "normal" cluster centroids. Those that are far enough from clusters are considered anomalous trajectories.

In [26] the authors propose a method in which trajectories are modelled as Hidden Markov Models (HMM) and grouped with hierarchical clustering. A similarity metric between HMMs is designed to determine the distance to clusters. A similar approach is taken in [27], which applies two-layers of hierarchical clustering to trajectories that are represented as a set of feature vectors that include location, velocity and size. The authors propose two similarity measures to detect point anomalies, by computing the probability of an anomaly when an object enters point k; and contextual anomalies, by computing the probability for an entire trajectory of being an anomaly.).

In [28], each trajectory is summarized as the parameters of a quadratic curve. At every spatial point, a Gaussian Mixture Model (GMM) is used to model the motion patterns of trajectories that pass through that point. For new observations, anomalies are detected as motions that display a low probability as predicted by the GMM model. The authors include appearance information by distinguishing between cars and pedestrians. However, the application of this technique is limited to constrained scenarios where the trajectories can be simplified as quadratic curves.

The authors in [23] generate a probability density function (PDF) from a Kernel Density Estimation (KDE) model for each pixel location in the image, taking object

location as well as size features. New observations are detected as anomalies if they have a low probability as predicted by the *pdf*. A similar approach is described by [29], in which object locations and transition times are employed to estimate a probability density function using a GMM model.

In [11], the authors extract motion related features (location, moving direction and velocity) and devise different strategies for point-anomaly detection, sequential anomaly detection (temporal context) and co-occurrence anomaly detection (spatial context). The first are addressed by computing histograms of features, and consider observations with low probability as anomalous. For sequential anomalies, the authors apply the data-mining CloSpan algorithm [30] to obtain the frequency of different sequences of feature vectors. For co-occurrence anomalies, HMM models are used.

These techniques are affected by challenging scenarios in which background subtraction and object tracking do not perform well. In particular, background subtraction performs poorly in crowded scenes, as well as situations with non-stationary backgrounds and sudden illumination changes. Object techniques often find difficulties in crowded scenarios in which occlusions are frequent, resulting in inaccurate tracks that impact the subsequent anomaly detection analysis negatively. Tracking does not scale well with object clutter, as it increases computational complexity and thus it is unsuitable for real-time applications.

2.7 Approaches based on pixel level abstractions

To overcome the limitations of techniques that extract motion features from trajectories, a number of methods have been recently proposed that do not use tracking and work at either the pixel or the region level by dividing the image in blocks. level. Some techniques do, however, incorporate features from the objects passing through pixel locations in the image, and therefore object segmentation is also required to extract these features.

In [18], the authors first slice the video in short sequences that are assumed to contain one event. This limits the applications of this approach and also makes it unable to locate the anomaly in space. After performing background subtraction, they compute a spatial histogram by blocks depicting object motion. Applying techniques from document-keyword clustering, the authors compute the co-occurrence of extracted features in video segments. Video segments that are sufficiently dissimilar from others are classified as anomalies.

In [13], a spatiotemporal video patch descriptor is computed for patches in the

image, which contains information from the spatial gradient at different spatial scales. A set of patches at different scales is extracted in the training phase to construct a database. For new observations, the authors propose to make use of a method to compose the patches of observed regions from patches in the database. If the new observations cannot be recomposed or if they can only be composed using the smallest patches, they are considered anomalous. Additionally, a strategy to progressively update the database of "normal" behavior is described.

The authors in [19] present a method to characterize the amount and location of motion inside a video segment (a collection of frames belonging to the same scene. These are described by proposed magnitudes "Total Motion" and "Average Motion", computed from the data obtained by background subtraction to spatially locate motion in video frames. Hierarchical clustering is then used to obtain the cluster centroids of normal events. For new observations, they are detected as anomalies if the distance to the closest cluster is above a threshold.

In [16], the authors extract two different pixel-wise features: pixel change frequency (number of transitions between foreground and background in a given time) and pixel change retainment (amount of time a pixel is considered as moving foreground). For noise reduction, these features are down sampled into an 8x8 super-pixel containing the average values. Similar to other approaches, the authors attempt to compute the posterior probability of an observation given past events. This is done in a Sequential Monte Carlo framework by modeling events as HMMs. Aside from point anomalies, this technique is able to detect contextual anomalies. However, this method requires substantial supervision as labeled instances of normal behavior are required.

The work proposed in[31] extracts texture information from patches in the non-stationary parts of the video. Patches are clustered into one of two behavior categories: A and B. Patches roughly correspond to moving blobs. Contextual information for each blob is extracted by taking into account the categories of the nearest blob neighbors. This approach shows good results on test sequences, however, the number of behavior categories or clusters is arbitrarily chosen for the application domain (bi-directional pedestrian motion in [31]) and may behave differently in other application domains.

The authors in [24] describe a Point-wise Motion Image in which motion information is coded in each color component (speed, orientation and motion duration, respectively). A correspondence measure is developed to detect anomalies in a given PMI. This technique is only capable of detecting point anomalies as it does not take into account sequential information.

Authors in [20] start from low level features from moving pixels: position and motion direction (from optical-flow). Quantized position and motion direction are assigned a word from a codebook. Unsupervised learning is employed, using techniques from language processing for clustering (Hierarchical Bayesian Models). Similar to other approaches, unseen observations are considered anomalies if they have a low likelihood. Additionally, anomalies based on interactions can be detected.

In [32], a framework for detecting different types of anomalies in video sequences is described. Pixel activity is model by a binary Markov chain that associates a feature vector (size, shape, motion) with the moving state. The transitions between moving and background states, along with the associated features, provide a statistical model for normal activity. For point-based anomalies, the authors take an approximation to the probability density function of normal activity from the model and classify unseen observations as anomalies based on low probability. For spatial co-occurring anomalies, Markov Random Fields (MRFs) are incorporated into the framework. Furthermore, the authors describe a framework for multicamera anomaly detection.

The aforementioned approaches pose significant advantages in scenarios in the presence of clutter, when compared to those that rely on tracking information. However, the proposed solutions are diverse in terms of contextual depth and the anomalies that can be detected, which makes it difficult to compare different techniques.

2.8 Evaluation methods

As mentioned in section 2.3, context and features determine the type of anomalies that different techniques are able to detect. Different solutions found in the literature target different anomalies, thus making it difficult to objectively compare approaches. In order to perform validation, an anomaly detection method must be tested on a data set of test sequences that contain instances of previously annotated anomalies. However, this process poses different challenges. Firstly, the concept of anomaly varies on each approach depending on the features extracted and whether or not contextual information is employed, which may even result in different anomalies being defined on the same datasets [8]. Therefore, establishing a ground truth is heavily reliant on subjective perception. Secondly, as mentioned in Section 2.4, the availability of anomalies in video datasets is scarce due to their infrequent nature, making it difficult to provide statistically significant performance metrics.

Due to these challenges, some authors have had to providing subjective performance assessments. In other works, authors manually annotate sequences from video datasets in order to provide performance metrics (precision/recall).

For approaches that consider the detection of anomalous trajectory paths, the work is simplified by annotating the ground truth of extracted paths in test sequences [29] [26] [23] [11]. However, there is no unified criteria on which paths are to be considered anomalous, and in some cases authors do not provide a criteria at all. In [29], ground truth is provided by three different subjects on the same set of trajectories. In approaches that perform clustering of trajectories, some authors [33] simply consider clear outliers are anomalies.

For approaches based on pixel-level abstractions, ground truth becomes more difficult to elaborate as it should label anomalous pixels in individual frames. In [34], the authors propose an evaluation framework for anomaly detection, consisting of ground-truth at different levels: frame level and pixel level. Frame level evaluation considers a correct detection if at least one anomalous pixel is found in a frame labeled as anomalous, without verifying the actual location of the anomaly. For pixel level evaluation, localised detections are compared to ground truth masks. A correct detection is considered if at least 40% of anomalous pixels are labeled correctly. Other authors ([35] [25]) have followed the same evaluation framework, and have been able to provide performance comparisons of different approaches on the same datasets, made public by [34].

2.9 Conclusions

Traditional pattern recognition approaches that perform event detection by modelling events of interest, are often domain specific and cannot be generalised for different applications. A recent paradigm shift towards anomaly detection consists on detecting anomalies by extracting the events in video sequences that differ from the surrounding spatio-temporal context. These approaches have a promising potential for their application in video-surveillance, in which it is often events that are "out of the ordinary" that human operators are more interested in detecting. In some cases, these events may have never been seen before.

However, anomaly detection poses challenges regarding how an anomaly is defined. While this has generated a great diversity of approaches, different techniques target different types of anomalies regarding context and and the features that are extracted, resulting in somewhat opposing definitions of anomalies. Additionally, the detection of more complex anomalies that involve complex spatio-temporal relationships between objects in the scene remains a challenge.

The diversity in existing techniques has also made it difficult to objectively compare the performance of different approaches. Apart from the difficulties derived

from considering anomalies from a subjective point of view, there exists no unified criteria on how to elaborate ground truth for existing datasets. All these factors, along with the the infrequent nature of anomalous events in video sequences, have made it difficult for authors to provide objective comparisons of existing approaches in the literature.

Chapter 3

Base system

3.1 Motivation

As discussed in 3, many authors perform anomaly detection by analysing the motion patterns extracted from single object trajectories. By employing object level abstractions, these techniques are capable of summarising activity in the scene in a period of time as a set of object trajectories. As previously mentioned, this makes it possible to construct models for normality in a fully unsupervised manner, by considering paths with low probability as anomalous. As an additional advantage, performance evaluation metrics can be obtained more easily, as ground truth can be constructed by manually labelling single trajectories rather than pixel regions in each frame.

However, object tracking poses significant challenges. Firstly, it is computationally intensive, which makes real-time execution difficult. Secondly, performance of current object tracking techniques is poor on scenarios with a high density of moving objects. Additionally, object occlusions can lead to inaccurate tracking. Tracking inaccuracies are potentially carried over to other stages of analysis, ultimately leading to a poor detection of anomalous events [32].

Approaches that work on pixel level abstractions are more suitable to handle cluttered scenarios, which remain an open challenge in many video surveillance applications [4]. Such is the work described by the authors in [22], which performs location-based statistical model of object attributes, rather than objects themselves.

The rest of this chapter is structured as follows. Section 3.2 describes the anomaly detection framework proposed by the authors in [22]. The remainder of the chapter detail the different analysis modules of the system.

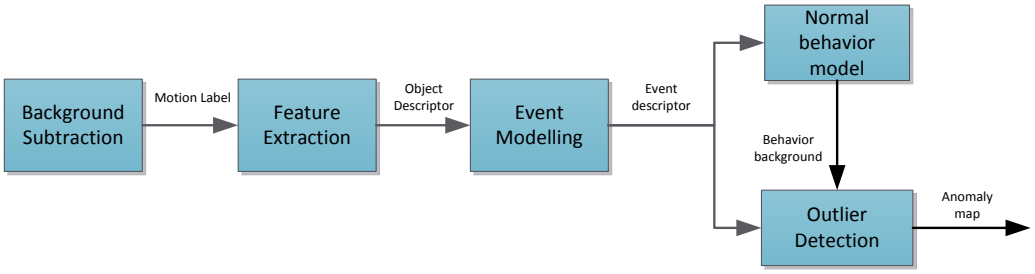


Figure 3.1: Overview of the base system

3.2 System overview

The authors describe a framework that is capable of modelling normal activity in the scene. For this purpose, a "background behaviour image" that captures background activity in the scene is constructed from training data. Activity in the scene is modelled at the pixel level by extracting features from regions in the image that are determined to be in motion. In order to detect anomalies, an image that captures current activity with associated features is constructed. Anomalous events are detected by comparing this image to the background behaviour image.

In this work, a fixed camera is assumed. Additionally, the authors impose the requirement of temporal stationarity of normal activity in the scene. Normal activity is defined as motion that is considered normal in the scene, which includes certain phenomena such as fluttering leaves in the background, moving water surfaces, or regular motion introduced by camera vibration.

An overview of the system is shown in Figure 3.1. At an initial stage, frames are captured from a static camera. For each frame, activity is characterised by labelling each pixel as either "moving" or "static". This *motion label image* can be computed employing existing background subtraction techniques [36].

In order to characterise the motion occurring at each pixel location, a pixel-level behaviour signature image is then computed. This behaviour signature consists on a *feature descriptor* that can include features such as the size, shape, speed and direction of objects passing through individual pixel locations.

In the event modelling stage, events modelled using a 2-state Markov chain. Events are defined as the behaviour signature (represented by the feature descriptor) left by moving objects over a time window. In this period of time, the pixel goes through transitions between the two aforementioned states, moving or static.

In the training phase, the event signatures of normal activity are employed to construct a *behaviour background image* that depicts normal activity in the scene. For anomaly detection, extracted events are compared against the behaviour background

to provide an anomaly map, depicting the location of anomalous motion.

3.3 Foreground segmentation

The purpose of the foreground segmentation module is to generate a binary mask that depicts the motion label of pixels in each frame. The labels can be either "moving" or "static". Based on a BackGround Subtraction (BGS) segmentation technique, a background model is created and then updated with incoming frames.

Let I_t denote a frame from a video sequence at time t, and $I_t(\vec{x})$ denote the pixel values at location $\vec{x}$; and let $L_t(\vec{x})$ denote the binary motion label for the same frame. To motion label is computed with background subtraction as follows:

$$L_t(\vec{x}) = \begin{cases} 1, & \text{if } |I_t(\vec{x}) - BG_t(\vec{x})| \geqslant \tau \\ 0, & \text{if } |I_t(\vec{x}) - BG_t(\vec{x})| < \tau \end{cases} \tag{3.1}$$

where $BG_t(\vec{x})$ is the background image for the current frame, and τ is a fixed threshold. For computational simplicity, the authors take a running Gaussian average [37] approach to computing the background image. The background model is updated progressively with each incoming frame in the following manner:

$$BG_t(\vec{x}) = \alpha I_t(\vec{x}) + (1 - \alpha) BG_{t-1}(\vec{x}) \tag{3.2}$$

where$\alpha$is the update coefficient and controls how quickly the background is updated with incoming frames.

3.4 Object size descriptor

The authors define events as the behavior signatures left by moving objects over time. The behavior signature is characterized by an object descritpor which embodies a feature or a set of features that describes that characteristics of the object passing through a given pixel location. These features could be appearance related (e.g., size, color, texture), or motion related (e.g., direction of movement, speed). For simplicity, the authors employ an object descriptor based on object size, due to its computational simplicity and because it has been found to perform well on different scenarios.

Let $F_t(\vec{x})$ denote the size descriptor. Given a an NxN pixel neighborhood centered around each pixel, the size descriptor can be computed as follows:

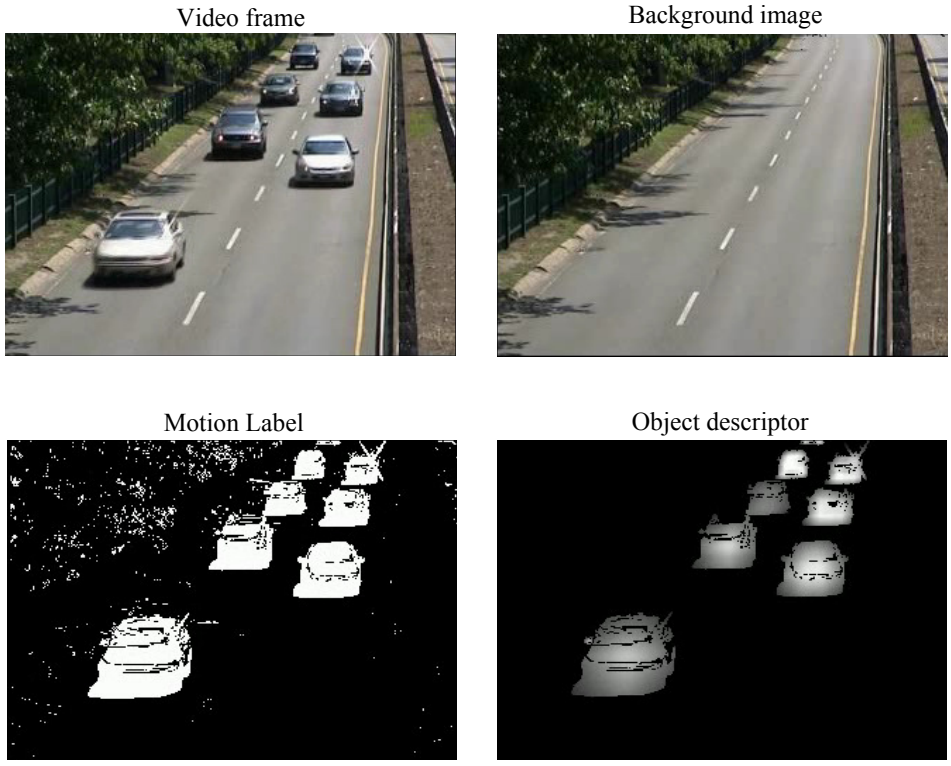


Figure 3.2: Example of Motion Label and Object Descriptor based on size for a video sequence

$$F_t(\vec{x}) = \frac{1}{N} \sum_{\vec{y} \in \mathcal{N}(\vec{x})} \delta(\vec{x}, \vec{y}) \tag{3.3}$$

where $\mathcal{N}(\vec{x})$ is the pixel neighborhood centered in $\vec{x}$, and $\delta(\vec{x}, \vec{y}) = 1$ if and only if locations $\vec{x}$ and $\vec{y}$ are both labeled as moving and belong to the same connected component in the motion label image, and zero otherwise. Connected component analysis [38] is employed to determine if two pixeles belong to the same connected component. By definition, the value of $F_t(\vec{x})$ is zero for those pixeles labeled static. For other locations, the descriptor has values greater than zero inside the object. Depending on the size of the neighborhood, the descriptor have increasing values until it saturates at 1 at the center, for big objects. For faster procesing, we discard the computation of the size descriptor for very small conneted components (e.g., spurious noise due to slight changes in illumination). An example of the size descriptor is shown in Figure 3.2.

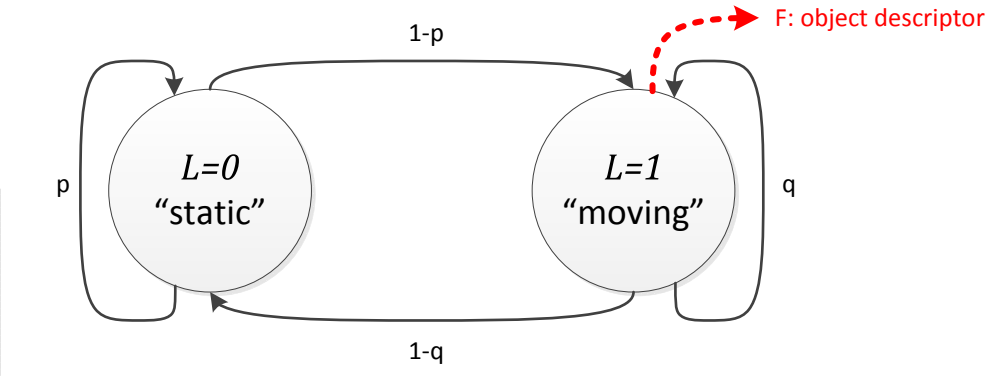


Figure 3.3: 2-state Markov chain model, with two possible states ("moving" and "static"). p and q are the state probabilities, and $1-p$, 1-q are the transition probabilities. [22]

3.5 Event modeling

Events are defined as the behavior signature (feature descriptor) that is left by a moving object over a period time, and modeled by a Markov transition model with two possible states. This model is shown in Figure 3.3. Based on this model, an event $E_t$is defined as a magnitude that is proportional to the joint probability of a particular sequence of state transitions over a w-frame time window, and the associated realization of the feature descriptor:

$$E_t(\vec{x}) = \sum_{k=t-w+1}^{t} (A_1 + A_3 F_t(\vec{x}))L_t(\vec{x})) + A_2 \mathcal{K}_t(\vec{x}) \tag{3.4}$$

where A_1, A_2 and A_3 are scalar constants, and $\mathcal{K}$ is a random variable that describes the number of state transitions that occur inside the time window. The authors presents their results with the following values for the constants $(A_1, A_2, A_3) = (0, 0, 1)$. Thus, the expression can be simplified as follows:

$$E_t(\vec{x}) = \sum_{k=t-w+1}^{t} F_t(\vec{x}) \tag{3.5}$$

since F_t only takes when non-zero values when L_t is non-zero as well. Events are therefore an accumulation in time of feature descriptors.

3.6 Anomaly detection framework

The authors a framework to perform detection of anomalous events with low computational costs. First, an image depicting the average activity in the scene is computed from a training sequence. Similar to background subtraction, this background image is then subtracted from the most recent event image in order to construct an anomaly map that displays where an anomaly has occurred.

3.6.1 Behavior background

Given a training sequence of M frames that only includes activity that is considered normal, a background behavior image B is defined as follows:

$$B(\vec{x}) = \max_{t \in [1,M]} \widetilde{E}_t(\vec{x}) \tag{3.6}$$

where $\widetilde{E}_t$ is the corresponding event descriptor for each frame. This abstraction captures the peak behavior signature in the training sequence in a single scalar per pixel, resulting in very low computational requirements. This implies that maximum activity during the training sequence will be considered normality. Unlike fully unsupervised methods, this approach requires to consider that the entire training sequence only contains normal activity. If any anomaly occurs during the training sequence, it is possible that it will be depicted in the behavior background, difficulting the detection of future anomalies in that region.

3.6.2 Behavior subtraction

In an analogous way to background subtraction, the detection of anomalous activity is reduced to thresholding the difference between observed events and the previously computed behavior background image, in the following way:

$$A_t(\vec{x}) = \begin{cases} anomalous, & \text{if } |E_t(\vec{x}) - B_t(\vec{x})| > \varphi \\ normal, & \text{if } |E_t(\vec{x}) - B_t(\vec{x})| \leq \varphi \end{cases} \tag{3.7}$$

where A_t is a binary anomaly map image, and φ is a fixed threshold.

Chapter 4

Analysis of Base System

4.1 Datasets

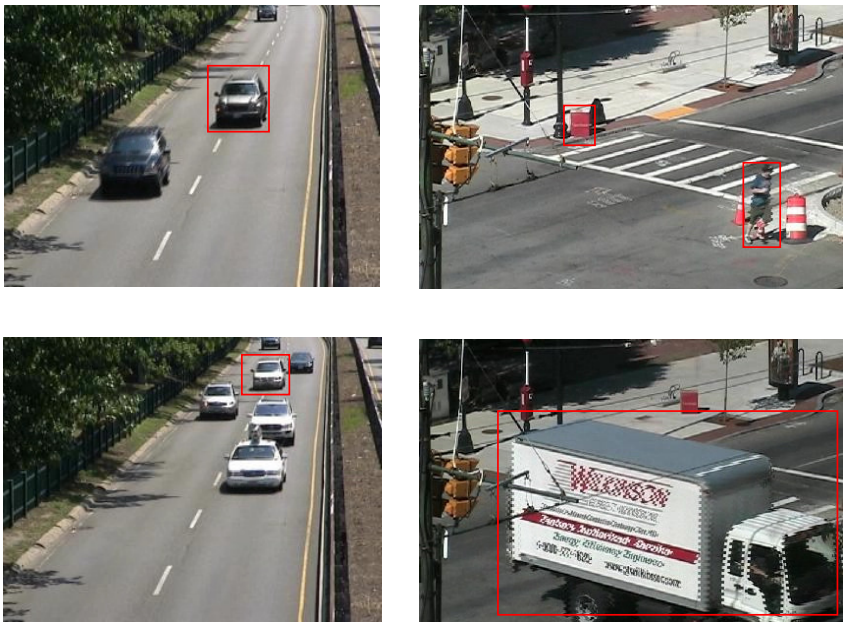
We have selected a number of video sequences from different scenarios containing instances of anomalous behaviour that can be of interest for video surveillance applications. Sequences include samples from the PETS 2006[1] , Changedetection.net[2] [39] and UCSD Anomaly Detection dataset[3] [34], available publicly. Additionally, we have also included videos recorded at the entrance hall of our building. The selected sequence cover three different scenarios: car traffic, indoor people transit, and outdoor people transit. Examples of common anomalies are explained in the following subsections.

4.1.1 Car traffic

Modelling normal behaviour of car transit can be easier to model due to the fact that car motion is spatially restricted to roads. In highways, it is expected that traffic flows in one direction at normal speeds. Common anomalies include any behaviour that deviates from this pattern: unusual speed (too fast, too slow), stalled vehicles, motion in opposite direction, motion in restricted areas (outside of road), and motion from objects other than cars, including pedestrians. Car overtaking can also be considered a spatial anomaly if we consider the crossing of lanes as unexpected behaviour. Vehicles of unusual sizes can also be anomalous if they are unexpected or restricted in a particular road.

On the other hand, road intersections can prove more difficult to model as expected behaviour is dependant on context, such as traffic status and traffic lights.

[1] http://www.cvg.rdg.ac.uk/PETS2006/data.html
[2] http://changedetection.net
[3] http://www.svcl.ucsd.edu/projects/anomaly/dataset.html



(a) (b)

Figure 4.1: (a) Examples of overtaking in a highway. (b) Examples of abandoned object by the sidewalk, pedestrian walking in restricted area, and big vehicle

Traffic lights make stationary cars normal behaviour intermittently. A robust anomaly detection system should be able to differentiate a stationary car is waiting at a red light, or is holding up normal traffic.

To evaluate the system in car traffic scenarios, two sequences have been selected from the Changedetection.net dataset. In the Highway sequence, there are two instances of car overtaking in a highway scenario. In the abandoned box sequence, there are instances of big vehicles and unusual pedestrian behaviour in a road intersection setting. Sample frames from these sequences are shown on Figure 4.1.

4.1.2 Indoor people transit

Anomalies in indoor scenarios are highly dependent on the context, as motion of individual subjects can sometimes be unpredictable. In general, stationary people are common where waiting is involved (airport lines, train platforms, seating areas). However, sometimes people display unusual behaviour even in areas with a high density of subjects. People running indoors (high speed), people loitering and large groups of people usually display motion that stands out from the surrounding context.

We have selected 6 sequences from the PETS 2006 dataset, which include instances of abandoned objects in a public transport station. While the detection of

Figure 4.2: Examples of people leaving unattended objects and loitering around the scene

abandoned objects has been widely studied in the literature [40][41], a robust anomaly detection system should be able to detect the unusual motion patterns of the subject that performs the action, especially if loitering is involved. Otherwise, inference from past information would be necessary to identify the potential subject [42][43].

Examples of anomalous behaviour from the selected sequences are shown in Figure 4.2 and Figure 4.3.

4.1.3 Outdoor people traffic

In an outdoor setting, we can expect anomalies of a similar nature than those indoor. In addition, certain vehicles are usually restricted in pedestrian areas, and can be considered anomalies (bikes, motorcycles, cars). People walking outside of footpaths or walking through restricted areas (such as fenced sections) are also common anomalies.

The UCSD Anomaly Detection dataset includes instances of people transit in a campus setting. In this particular scenario, anomalies are related to unusual speed (bikes, skates) and unusually large moving objects objects (small carts). Examples of anomalies from this dataset are shown in Figure 4.4.

4.2 Evaluation of base system

4.2.1 Implementation

The base system has been implemented in C++ using the OpenCV computer vision library[4].

[4] http://www.opencv.org

(a) (b)

Figure 4.3: (a) Example of abandoned objects and nearby suspects. (b) Example of abnormally large objects in motion

Figure 4.4: Sample anomalous events from UCSD Anomaly Detection Dataset

4.2.2 Evaluation procedure

In order to evaluate the performance of the base system, we first produce the behavior background for each video from a training sequence. The duration of each training sequence if at least 1000 frames. For videos from UCSD Anomaly Detection Dataset or Entrance Hall videos, separate training sequences are provided. For the rest of the videos, a training sequence of frames is selected from each video. These training sequences are carefully selected in order to avoid anomalous behaviour to become part of the behavior background.

For detection, we employ those previously computed behavior background images to compute the anomaly map for each frame in the video sequences, and evaluate whether or not the previously described anomalies are successfully detected.

4.2.3 Behavior background computation

As previously mentioned, inclusion of anomalous behavior in the training phase will result in this behavior to be expressed in the background. However, in those scenarios in which stationary objects or subjects are common, these will appear in the behavior background as by definition, as the event descriptors accumulate over the temporal window. This is particularly relevant in scenarios where stationary people are common, or near intersections in which it is normal for cars to be waiting at traffic lights. In practice, anomaly detection in this area will be difficult if the anomalous motion occurrs in the same location as the stationary objects in the behavior background. Furthermore, similar accumulation of stationary masks can occurr if the background model is not correctly initializated at the background subtraction stage.

Examples of this phenomenon are shown in row (2) of Figure 4.5. For the indoor sequence (left), the behavior background displays a person that remains stationary for most of the duration of the video. For the intersection sequence (right), a stationary bike and car that remain stationary due to a red traffic light are shown in the behavior background.

4.2.4 Successful detections

Since the algorithm is based on a size feature descriptor, anomalies are detected when objects have an unusually large size at an unusual location. We can distinguish between three types of anomalies that can be successfully detected with this algorithm: motion in restricted or unusual locations, stationary objects, and objects with unusually large size. In the first case, any motion that occurrs in areas in which motion is not normal will be shown in the anomaly map. For stationary objects, the

Figure 4.5: Examples of trained Behavior Background images. Correct (1) and problematic (2)

event descriptor will simply accumulate the size descriptor in time, so these region will be highlighted after the behavior background subtraction. This is also the case for unusually slow objects. Finally, objects of larger sizes than usual will also be detected, as they have higher values of the size descriptor.

Examples of these cases of successful detections are shown in Figures 4.6, 4.7 and 4.8. In Figure 4.6, car overtaking is successfully retrieved in a highway scenario, due to the fact that the car crosses between lanes, and the behavior background does not display regular motion in the area between the two lanes. In Figure 4.7, the anomaly map displays only the anomalous static regions, while ignoring other motion in the scene. In Figure 4.8, we see an example of a big vehicle being successfully detected as an anomaly.

4.2.5 Unsuccessful detections

Other types of anomalies cannot be successfully detected by the algorithm due to the limitations of the size descriptor. In particular, we have identified the following anomalies that fail to be detected:

1. Motion at unusually high speeds. Unless the objects are too large, if they are moving too fast when compared to the surrounding context, the system fails to detect them. As explained in section 3.5, the event model accumulates the values of the size descriptor over a temporal window. When objects move too fast, the object descriptor does not accumulate in the same position, and therefore the event signature of that object does not stand out from surrounding motion. An example of this is shown in Figure 4.9 (left), where a person is seen skating in the same area as pedestrians at a significantly higher speed, but fails to be detected as an anomaly.

2. Motion in opposite or unusual directions. When unusual motion occurrs in

Figure 4.6: Detection of a car overtaking as an anomaly

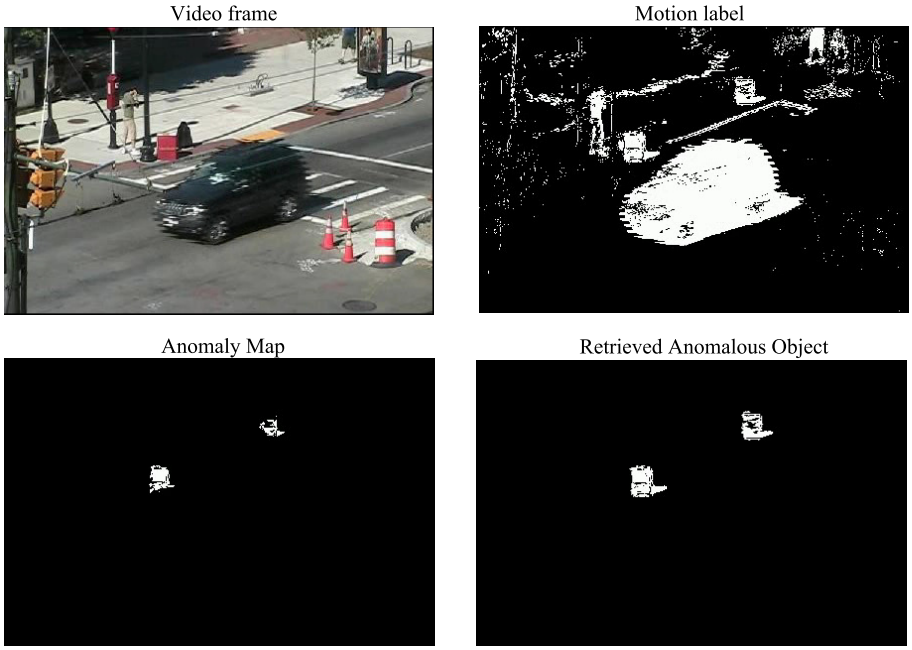


Figure 4.7: Successful detection of stationary objects (box is removed from original location)

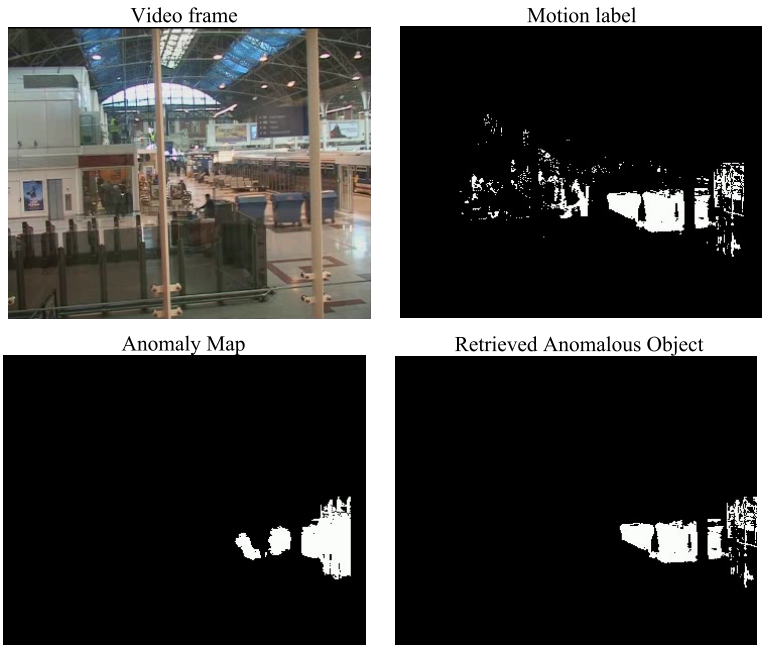


Figure 4.8: Successful detection of large moving object as an anomaly

the same locations as regular motion, but in opposite directions, the system is incapable of highlighting anomalies, as motion direction is not taken into account. An example is shown in Figure 4.9 (right), where a pedestrian crosses the street in an unusual direction, but fails to be detected because his behaviour signature does not stand out in the behavior background, which displays paths normally followed by cars in the same location.

3. Stationary objects of small size. Equation 3.3 in Chapter 3 defines the size descriptor as the average number of pixels from a connected component in the motion label image inside a pixel window. However, the size of the pixel window is fixed and may result in behavior signatures near zero for very small objects that may still be of interest. This highlights the fact that a fixed spatial window size is suboptimal in those scenarios in which objects are shown at different sizes due to perspective and camera resolution. An example of a missed detection of a small abandoned object is shown in Figure 4.10.

Figure 4.9: Difficult anomalies related to motion characteristics: speed (left) and moving direction (right)

Figure 4.10: Example of a missed detection due to small size of stationary object

Chapter 5

Proposed Enhancements

5.1 Introduction

As discussed in Chapter 3, the base system presents limitations when dealing with anomalies that are related to unusual motion parameters (speed or direction of movement), as well as the dependence of the size descriptor on screen resolution and actual object size.

In order to address these issues, we propose a modified size descriptor that is independent of video frame resolution and object size. This way, anomalous behaviour displayed by very small objects is correctly displayed on the anomaly map. Additionally, we build on the proposed framework by the authors in [22] to handle a vector of motion-based features.

The remainder of this chapter is structured as follows. In section 5.2, we describe the proposed modified size descriptor. In section 5.3, we describe the approach to incoporate motion features to the model.

5.2 Resolution independent object size descriptor

As we recall from Eq. 3.3, the object descriptor saturates at a value of 1 for objects that are considerably larger than the spatial window centred around each pixel. An example of a size descriptor image computed for a sample motion label image is depicted in Figure 5.1, where a small object due to perspective will leave a very small behavior signature. The chosen size for the spatial window has an impact on the types of anomalies that can be detected. For very small window sizes, the event model will accumulate motion labels in time, and any motion that occurs outside of normal areas of the behaviour background will be considered anomalies. For window

Figure 5.1: Size descriptor image (right) computed for corresponding motion label (left)

sizes that are significantly larger than the average object size, only very large objects will stand out from the behaviour background, as well as large stationary objects.

Therefore, the spatial window size has to be carefully chosen depending on the average size of objects in the scene, as well as image resolution. For many applications, this is unfeasible if there are objects from different sizes (people, vehicles, luggage) or size variations due to camera perspective.

In order to account for these expected variations in size, we propose an object size descriptor with a variable window size that depends on object size. The proposed size descriptor is defined as follows:

$$F_t(\vec{x}) = \frac{1}{N(\vec{x})} \sum_{\vec{y} \in \mathcal{N}(\vec{x})} \delta(\vec{x}, \vec{y}) \tag{5.1}$$

where $N(\vec{x})$ is the variable window size, and defined as follows:

$$N(\vec{x}) = \min(5, \sigma * S(\vec{x})) \tag{5.2}$$

where $S(\vec{x})$ describes the number of pixels in the connected component to which $\vec{x}$ belongs, and σ is an arbitrary coefficient. A comparison between the resulting size descriptor from the original method (left) and the proposed method (right) is shown in Figure 5.2.

5.3 Vector behavior subtraction with motion features

5.3.1 Event model and behaviour background

The authors in [22] describe a framework to accommodate multiple arbitrary features based on a vector feature descriptor $\vec{F}$. A suboptimal approach to the joint proba-

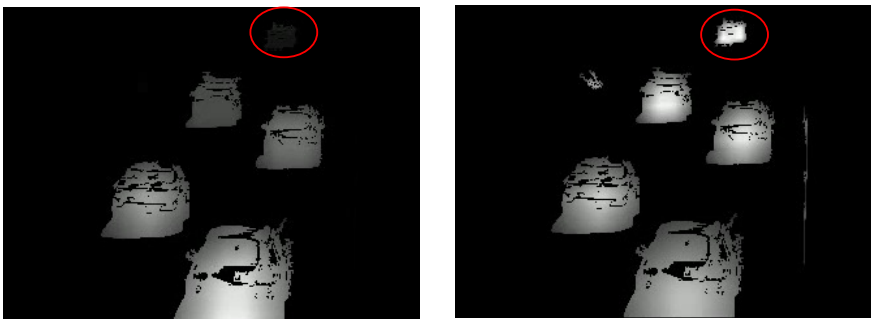


Figure 5.2: Comparison between size descriptors with original method (left) and proposed method (right)

bility of a realisation of state transitions of a w-frame time window is described, by computing events independently for each vector component as follows [22]:

$$E_t^i(\vec{x}) = \sum_{k=t-w+1}^{t} F_t^i(\vec{x}) \tag{5.3}$$

where $F_k^i(\vec{x})$,$i = 1, 2, \ldots n$ is the i-th component of the vector descriptor, and $E_t^i(\vec{x})$ is the event model for the i-th component at instant t.

Likewise, the behaviour background image is computed independently for each component as follows [22]:

$$B^i(\vec{x}) \max_{t\in[1,M]} \widetilde{E_t}^i(\vec{x}) \tag{5.4}$$

where $B^1(\vec{x}) \ldots B^n(\vec{x})$ are the behaviour backgrounds for each of the n components, M is the duration of the training sequence and $\widetilde{E_t}^i(\vec{x})$ are the event models for each component at instant t in the training sequence. In the detection phase, behaviour background subtraction is then applied as described in Eq. 3.7 to compute the anomaly map for each component.

5.3.2 Motion feature extraction

As discussed in the previous chapter, one of the limitations of employing an object descriptor that relies only on size, is the inability do detect abnormal behaviour caused by unusual motion parameters, such as motion direction or speed. The authors [22] propose a 5-component Feature vector that describes the size of an object (size descriptor value) passing through a location $\vec{x}$ in 4 directions (leftwards, rightwards, upwards and downwards), as well as an additional component for stationary

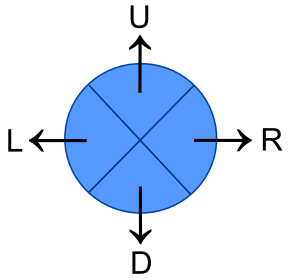


Figure 5.3: Four directions considered: Upward (U), Downward (D), Leftward (L), Rightward (R)

behaviour. This approach equates to having considering 5 different event models for each direction independently.

In order to determine which direction an object is moving, we first compute the values of optical flow vectors for each pixel. Between two consecutive frames, optical flow vectors depict the distribution of apparent velocities of intensity patterns in the image [44]. Given the values of optical flow vectors$\vec{u}_{\vec{x}}$ for each pixel in the current frame, the dominant motion vector $\vec{v}_{\vec{x}}$ is computed as follows:

$$\vec{v}_{\vec{x}} = \frac{1}{S} \sum_{\vec{y} \in \mathcal{S}(\vec{x})} u\vec{y} \tag{5.5}$$

where $\mathcal{S}(\vec{x})$ is the connected component to which$\vec{x}$ belongs, and S is the number of pixels in the same connected component. The angle φ of vector $\vec{v}_x = (x, y)$ is computed as follows:

$$\varphi = \arctan\left(\frac{y}{x}\right) \tag{5.6}$$

The dominant direction of movement is then determined by thresholding angle φ according to the diagram displayed in Figure 5.3.

Chapter 6

Experimental work

6.1 Introduction

In order to validate the proposed modifcations, we evaluate the performance of the enhanced system on those scenarios in which the base system displayed problematic detections. In section 6.2, we discuss the performance of the system with the proposed modified size descriptor. Section 6.3 discusses potential applications of detection of anomalies based on motion direction features.

6.2 Modified size descriptor

As discussed in 3, the fixed spatial window in the size descriptor difficults the detection of anomalies caused by very small objects in the scene. The proposed method in Chapter 5, with a variable spatial window that is invariant to scale, attempts to mitigate this problem while maintaing the correct detection for the rest of anomalies. In Figures 6.1 and 6.2, two cases of successful detections of very small unattended objects are shown, that go undetected in the original system. Anomaly maps for base system and modified system are displayed.

This approach, however, displays erratic detections when the motion label is affected by noise due to sudden changes in illumination. This situation tends to group different moving objects into a single large blob, which due to its size it is very likely to cause noise in the anomaly map. As can be seen in Figure 6.3, a transient illumination change causes an incorrect motion label, which results in noisy artifacts in the anomaly mask, which correspond to large blobs in the motion label. This phenomenon is more sever if the illumination change persists in time due to slow background updates.

Figure 6.1: Detection of small unattended object. Anomaly map for base system (A) and with proposed modified size descriptor (B)

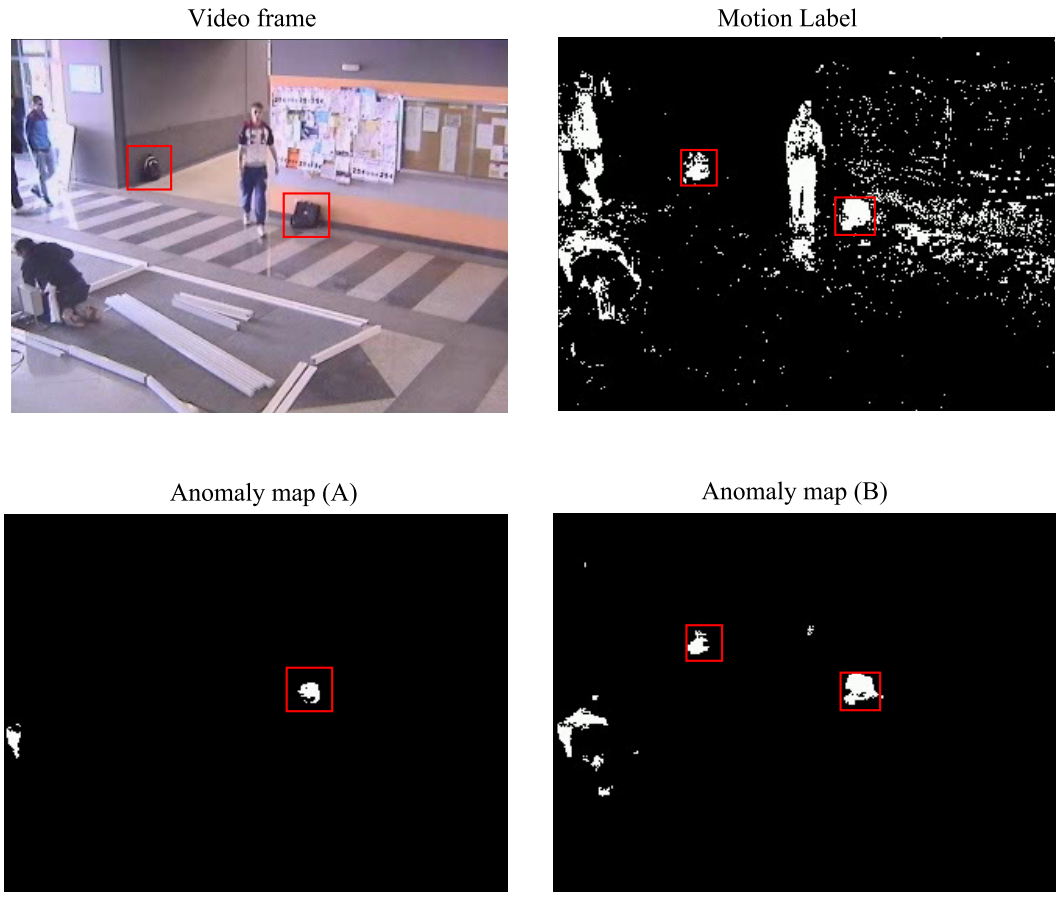


Figure 6.2: Detection of unattended objects. Anomaly map for base system (A), where the smallest object is not detected, and with proposed modified size descriptor (B)

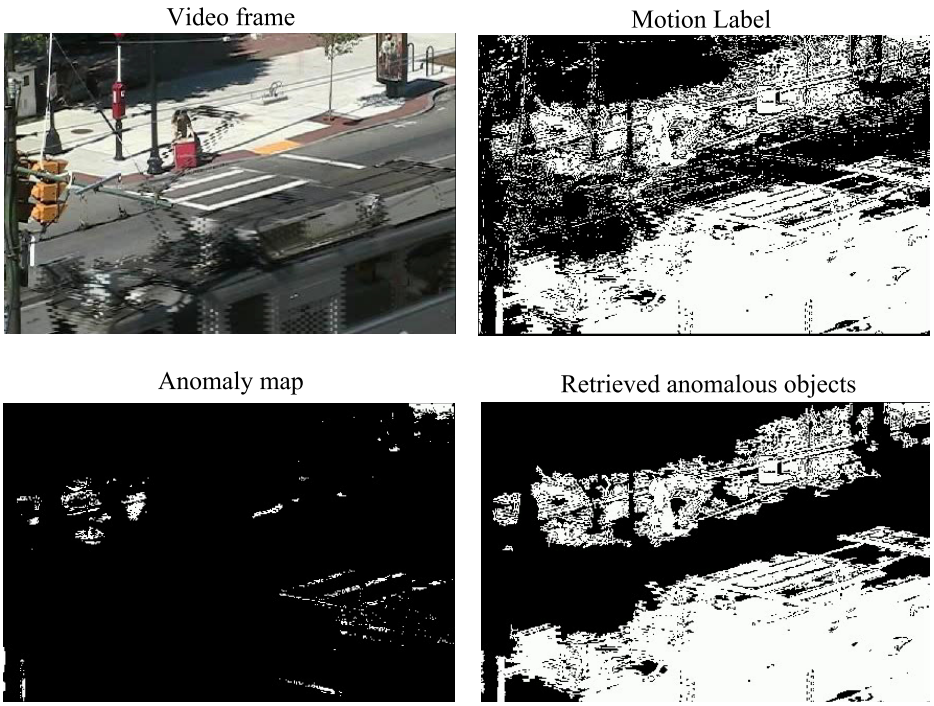


Figure 6.3: Example of the effects of an incorrect motion label due to sudden illuminationc hanges

6.3 Vector behavior subtraction with motion features

6.3.1 Behavior background

As described in Chapter 5, now we have a behavior background image for each component in the feature descriptor vector. In the described approach, we consider a behavior background image for each direction of motion (upward, downward, leftward, rightward), plus an additional one for static objects. These behavior backgrounds now display behavior from objects that display dominant motion in each particular direction, as opposed to motion from all directions. This allows for the detection of anomalies due to unusual moving direction, in areas in which this behavior would not be anomalous by considering size features alone. An example of directional behavior backgrounds is shown in Figure 6.4.

6.3.2 Anomaly detection

Anomalies are detected independently in each of the considered moving directions. The directional behavior backgrounds indicate the areas in the scene in which each direction is dominant. Therefore, the system is now able to detect anomalies caused

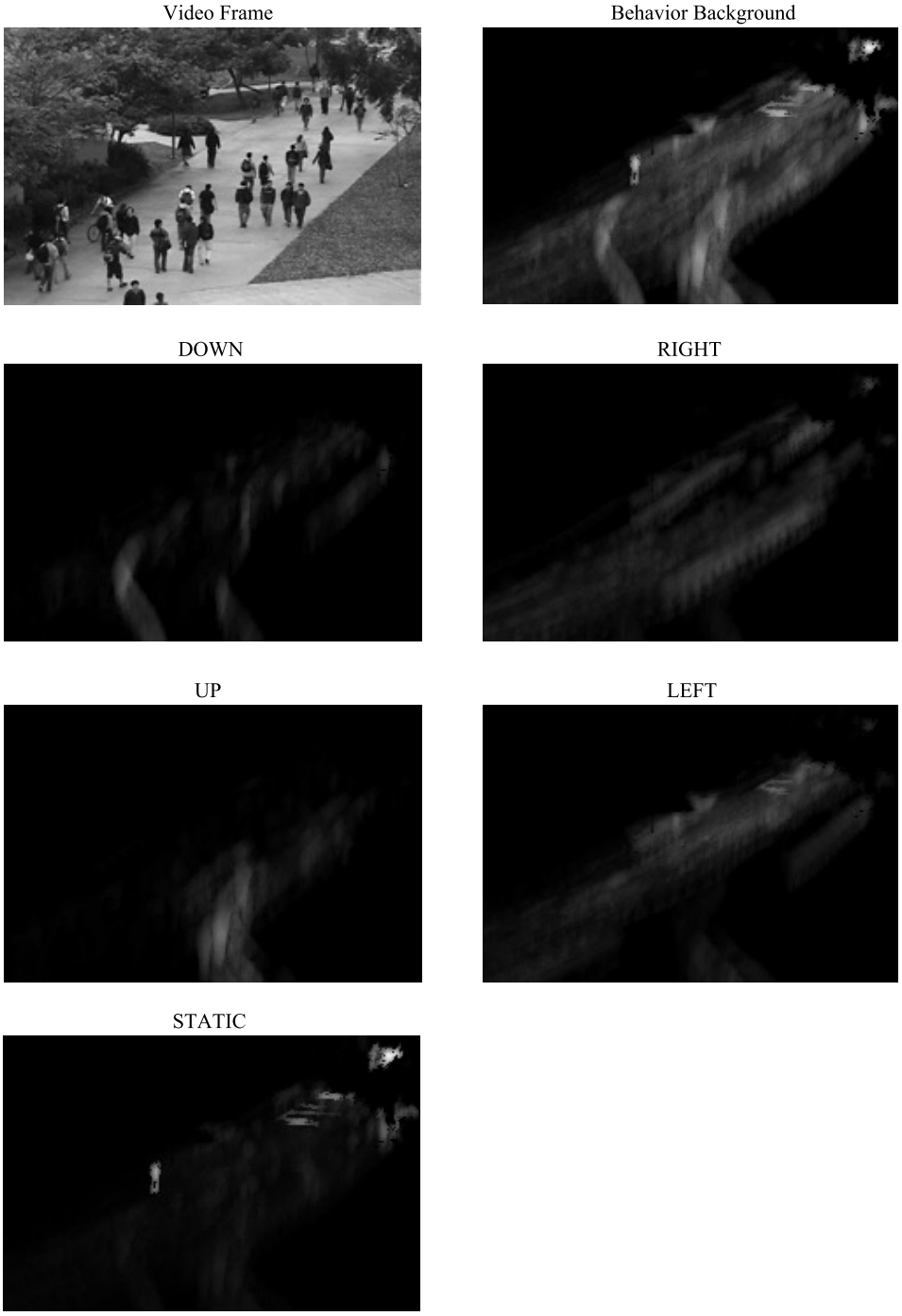


Figure 6.4: Example of behavior background computed with base system (first row), and for dominant directions of motion

by objects that move in an area in which the direction they are taking is not dominant. This has obvious applications in scenarios in traffic scenarios for the detection of subjects moving opposite incoming traffic. For indoor people transit scenarios, we have found that this approach aids in highlighting subjects moving in unusual directions that would otherwise go unnoticed. An example is shown in Figure 6.5. In this scene, a subject is slowly moving rightwards and eventually leaves a piece of luggage unattended. Relying on the size descriptor alone, the system is able to successfully highlight the object once it has already been left unattended. The improved system, however, is able to highlight the suspect moving towards the right direction with the object, *before* it is left unattended, as this direction of motion is unusual in that particular area of the scene.

6.4 Conclusions

In this chapter, we discuss the performance of the proposed enhancements in scenarios that pose challenges to the base system. The modified size descriptor allows the system to detect anomalous motion of very small objects, especially small unattended objects. For the behavior subtraction framework with vector features, successful applications in the detection of suspicious behavior have been shown.

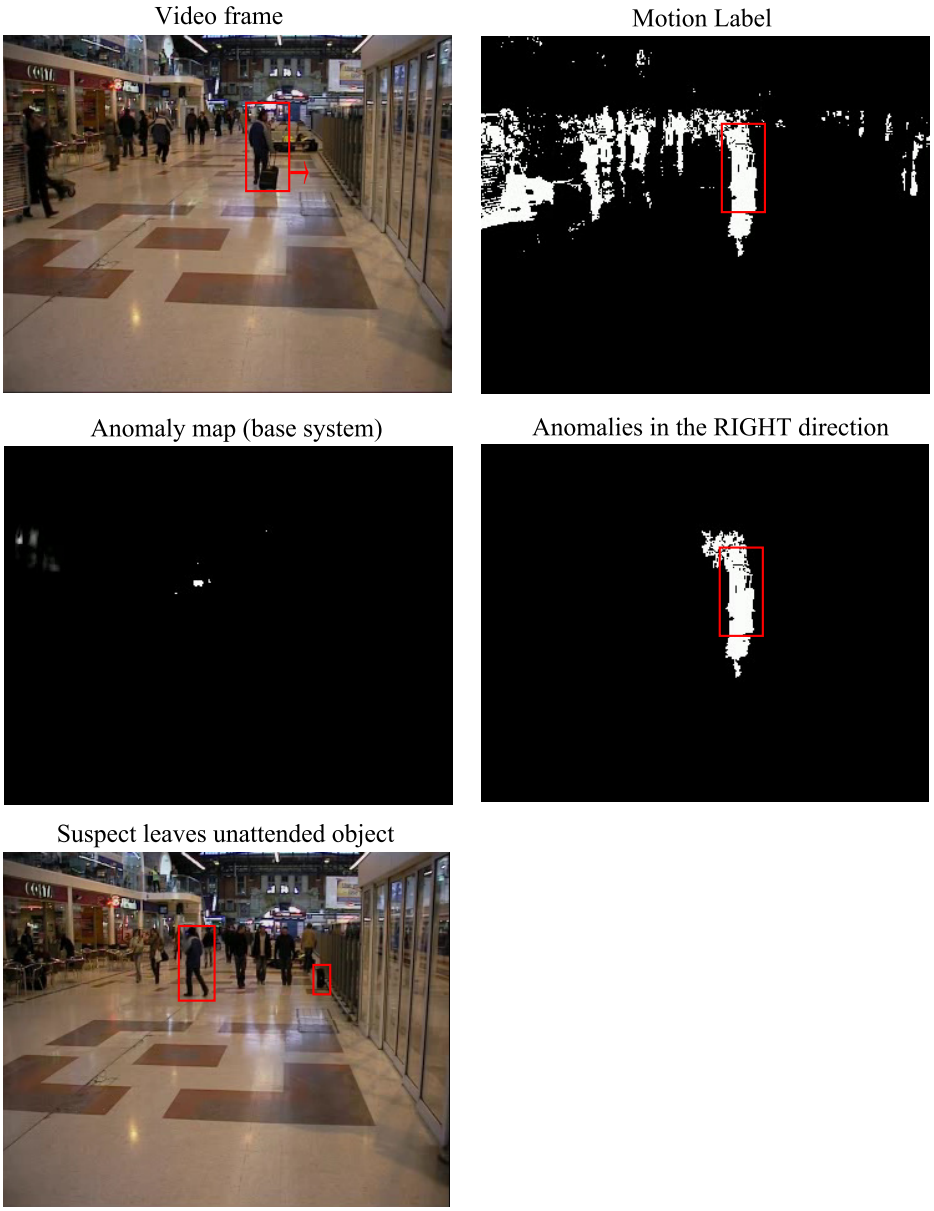


Figure 6.5: Comparison of detection of subject moving in unusual direction

Chapter 7

Conclusions and future work

7.1 Summary of work

In this project, a comprehensive study of an existing anomaly detection framework has been carried out. The detection of anomalies in video surveillance sequences has gathered considerable interest as a research topic in the recent years. Traditionally, researchers have taken a pattern recognition approach to detect a set of previously defined events. However, these approaches are generally limited to constrained scenarios and can not be easily generalised for arbitrary behaviour. More recently, there has been a paradigm shift towards statistically modelling normal behaviour in a scene, focusing on detecting behaviour that stands out from the surrounding context. The contributions of this work can be summarised as follows:

1. Identification of current challenges in anomaly detection. An extensive study of the State of the Art has been carried out to identify key challenges in anomaly detection. In order to compare existing approaches, three key areas have been identified: definition of anomaly, extracted features, and evaluation method.

2. Implementation and evaluation of an existing framework for anomaly detection. An existing approach from the literature has been selected and implemented. This technique models behaviour in the scene with pixel-based abstractions rather than object-based abstraction, making it more suitable for scenarios with higher density of objects. A set of video sequences containing anomalies from common surveillance scenarios have been selected from publicly available datasets. The system has been evaluated on these video sequences in order to identify possible shortcomings.

3. Proposal of improvements to the base algorithm for challenging scenarios. A

modified size descriptor that is invariant to spatial scale has been proposed in order to improve the system for the detection of anomalous behaviour of small objects. Additionally, motion features have been included in order to detect anomalies caused by object motion in irregular directions.

7.2 Conclusions

Anomalies can be broadly defined as an observation that stands out from the surrounding context. While intuitive, this definition has lead to subjective interpretations of anomalies, which have resulted in very diverse approaches aimed at solving the same problem. Depending on the nature of the features extracted to model normal behaviour and anomalies, different anomalies can be considered. This has made it difficult to compare existing techniques, as they often rely on different definitions for anomalies, and can result in different authors identifying different anomalies on the same datasets. Additionally, the infrequent nature of anomalous events makes then infrequent in current video datasets, and most authors evaluate their approaches on a very limited number of video sequences.

Existing approaches can be broadly classified between those that extract information from object tracking, and those that rely on pixel level abstractions. While the former allow for simpler modelling of normal behaviour, they are unsuitable in real scenarios due to potential inaccuracies in object tracking. Furthermore, object tracking techniques are computationally intensive and cannot be applied in real-time situations. Pixel level abstractions, such as the approach selected for its implementation, have proven more robust in scenarios with a higher density of objects. The selected approach performs modelling directly on pixel matrices, which allows for very fast computation.

The base system has shown to perform well on different scenarios for the detection of spatial anomalies. The proposed modifications have shown to solve some of the shortcomings of the original system. However, detection of objects displaying anomalous speeds, or more complex context-related anomalies, remain an open issue.

7.3 Future Work

In this project, we have identified different areas that remain an open issue in the field of anomaly detection in video surveillance. The following lines of research can be considered:

- Elaboration of a comprehensive dataset for evaluation and validation of anomaly

detection. A robust evaluation framework should be able to accommodate the subjective nature of anomalies in existing approaches. Furthermore, longer video sequences should be included to be able to properly model normal behaviour.

- Inclusion of speed-related features in the base system. The base system is able to accommodate a vector of object descriptor features, of which motion direction has already been considered. This framework can be extended to include speed-related features, in order to make the system robust against anomalies due to unusual speeds.

- Elaboration of strategies to provide robustness against changing context. Most approaches in the field work under the assumption of a "stationary normality". This simplification is often unrealistic, as motion patterns of subjects and objects is often conditioned on a context that can change in time. A robust anomaly detection system should be able to adapt to changing contexts. For the base system, a way to update the behaviour background in an online manner should be explored.

Bibliography

[1] M. Green, J. Reno, R. Fisher, and L. Robinson, "The appropriate and effective use of security technologies in U.S. schools: A guide for schools and law enforcement agencies series: Research report," *National Institute of Justice*, 1999.

[2] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti, "Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking," *IEEE Signal Process. Mag.*, vol. 22, no. 2, pp. 38–51, 2005.

[3] C. S. Regazzoni, A. Cavallaro, Y. Wu, J. Konrad, and A. Hampapur, "Video Analytics for Surveillance: Theory and Practice," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 16–17, 2010.

[4] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *The Visual Computer*, 2012.

[5] S.-W. Joo and R. Chellappa, "Attribute Grammar-Based Event Recognition and Anomaly Detection," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pp. 107–107.

[6] J. C. San Miguel and J. M. Martinez, "Robust unattended and stolen object detection by fusing simple algorithms," in *Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on*, pp. 18–25, IEEE, 2008.

[7] J. C. SanMiguel, M. Escudero-Vinolo, J. M. Martinez, and J. Bescós, "Real-time single-view video event recognition in controlled environments," in *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, pp. 91–96, IEEE, 2011.

[8] O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 865–878, 2012.

[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, pp. 1–58, July 2009.

[10] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3313–3320, 2011.

[11] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, pp. 323–333, Mar. 2011.

[12] T. Xiang and S. Gong, "Video behaviour profiling and abnormality detection without manual labelling," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1238–1245 Vol. 2, IEEE, 2005.

[13] O. Boiman and M. Irani, "Detecting Irregularities in Images and in Video," *Int J Comput Vision*, vol. 74, no. 1, pp. 17–31, 2007.

[14] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.

[15] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas, "Abnormal detection using interaction energy potentials," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3161–3167, 2011.

[16] P. Cui, L.-F. Sun, Z.-Q. Liu, and S.-Q. Yang, "A Sequential Monte Carlo Approach to Anomaly Detection in Tracking Visual Events," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–8, 2007.

[17] T. Xiang and S. Gong, "Video Behavior Profiling for Anomaly Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 5, pp. 893–908, 2008.

[18] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, pp. II–819–II–826 Vol.2, 2004.

[19] J. Oh, J. Lee, and S. Kote, "Real Time Video Data Mining for Surveillance Video Streams," in *Knowledge Discovery and Data Mining*, pp. 222–233, Berlin, Heidelberg: Springer Berlin Heidelberg, Apr. 2003.

[20] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 3, pp. 539–555, 2009.

[21] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pp. 55–61, 2011.

[22] P. M. Jodoin, V. Saligrama, and J. Konrad, "Behavior Subtraction," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4244–4255, 2012.

[23] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE, 2008.

[24] Q. Dong, Y. Wu, and Z. Hu, "Pointwise Motion Image (PMI): A Novel Motion Representation and Its Applications to Abnormality Detection and Behavior Recognition," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 19, no. 3, pp. 407–416, 2009.

[25] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2112–2119, IEEE, 2012.

[26] F. Jiang, Y. Wu, and A. K. Katsaggelos, "A Dynamic Hierarchical Clustering Method for Trajectory-Based Unusual Video Event Detection," *Image Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 907–913, 2009.

[27] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 9, pp. 1450–1464, 2006.

[28] T. Zhang, H. Lu, and S. Z. Li, "Learning semantic scene models by object classification and trajectory clustering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1940–1947, 2009.

[29] I. Saleemi, K. Shafique, and M. Shah, "Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 8, pp. 1472–1485, 2009.

[30] X. Yan, J. Han, and R. Afshar, "CloSpan: Mining Closed Sequential Patterns in Large Datasets," *Proceedings of the Third Siam International Conference on Data Mining*, 2003.

[31] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Detecting contextual anomalies of crowd motion in surveillance video," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 1117–1120, 2009.

[32] V. Saligrama, J. Konrad, and P. Jodoin, "Video Anomaly Identification," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 18–33, 2010.

[33] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-Based Anomalous Event Detection," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1544–1554, 2008.

[34] V. Mahadevan, W. Li, V. Bhalodia, N. C. V. Vasconcelos, and P. R. C. . I. C. on, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010.

[35] B. Antic and B. Ommer, "Video parsing for abnormality detection," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2415–2422, 2011.

[36] M. Piccardi, "Background subtraction techniques: a review," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, pp. 3099–3104, 2004.

[37] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 780–785, 1997.

[38] L. Di Stefano and A. Bulgarelli, "A simple and efficient connected components labeling algorithm," in *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pp. 322–327, 1999.

[39] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pp. 1–8, 2012.

[40] H. Kong, J. Y. Audibert, and J. Ponce, "Detecting Abandoned Objects With a Moving Camera," *Image Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2201–2210, 2010.

[41] Y. Tian, R. S. Feris, H. Liu, A. Hampapur, and M.-T. Sun, "Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, no. 5, pp. 565–576, 2011.

[42] M. Bhargava, C.-C. Chen, M. S. Ryoo, and J. K. Aggarwal, "Detection of abandoned objects in crowded environments," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pp. 271–276, 2007.

[43] M. Bhargava, C.-C. Chen, M. S. Ryoo, and J. K. Aggarwal, "Detection of object abandonment using temporal logic," *Machine Vision and Applications*, vol. 20, pp. 271–281, June 2009.

[44] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, Aug. 1981.